a) Existing Methods (✘ Global Alignment)

**Contrastive Loss**

Binary Focal Loss

Real

Fake

Audio

Audio    Similarity    Video

1

1

Video

1
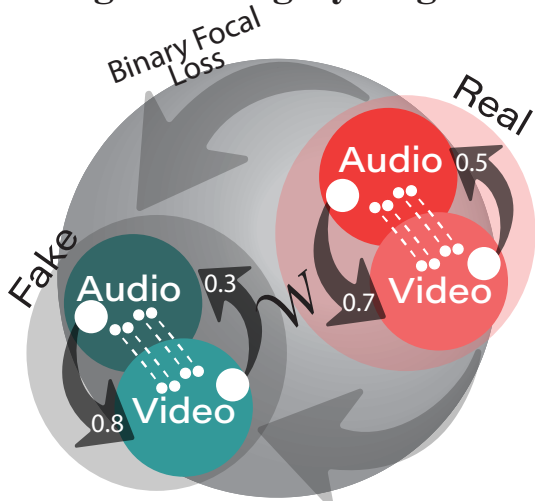
**Heavy + Modality Dependent**

Audio → Audio Transformer

**Modality-dependent**

Video → Video Transformer

Fusion

Real or Fake
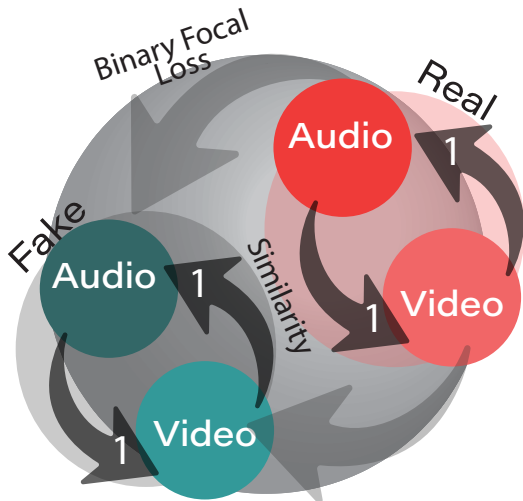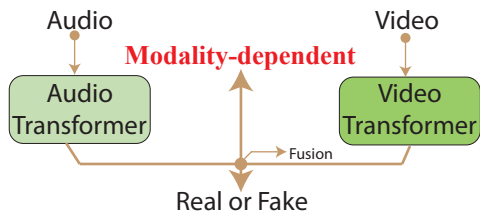
b) HashFormer (Ours) (✔ Fine-grained Alignment)

**Weighted Forgery Alignment**

Binary Focal Loss

Real

Fake

Audio    0.5

Audio    0.3    W    0.7    Video

0.8    Video

**Light Weight Modality Agnostic**

Audio •————— Video ————• Landmark based distillation

**Missing ?** → SFA → **Modality-Agnostic**

HashFormer

Real or Fake